Flexible Covariance-Targeting Volatility Models Using Rotated Returns

DIAA NOURELDIN
Department of Economics, University of Oxford,
& Oxford-Man Institute,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK
diaa.noureldin@economics.ox.ac.uk

NEIL SHEPHARD
Department of Economics, University of Oxford,
& Oxford-Man Institute,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK
neil.shephard@economics.ox.ac.uk

KEVIN SHEPPARD
Department of Economics, University of Oxford,
& Oxford-Man Institute,
Eagle House, Walton Well Road, Oxford OX2 6ED, UK
kevin.sheppard@economics.ox.ac.uk

July 26, 2011

Abstract

This paper introduces a new class of multivariate volatility models which is easy to estimate using covariance targeting. The basic structure is to rotate the returns and then to fit them using a BEKK model of the time-varying covariance whose long-run covariance is the identity matrix. The extension to DCC type models is given, enriching this class. Inference for these models is computationally attractive, and the asymptotics is standard. The techniques are illustrated using recent data on the S&P 500 ETF and some DJIA stocks.

Keywords: DCC; GARCH; orthogonal GARCH; multivariate volatility; diagonal models; common persistence; covariance targeting; predictive likelihood.

JEL classification: C32; C52; C58.

1 Introduction

Search is still ongoing for multivariate volatility models with flexible dynamics and ease of application in moderately large dimensions. Modelling and forecasting multivariate volatility is not only crucial for asset pricing and optimal portfolio allocation, but also to characterise the systemic risk profile of individual firms. **?** illustrate the importance of modelling and forecasting the conditional covariance matrix of asset returns, where they show that a rise in a stock's return volatility and correlation with the market magnifies its contribution to their proposed measure of systemic risk. Highly leveraged financial companies in the recent financial crisis are a case in point.

The crisis forcefully demonstrated the need for more robust models to capture and project financial risk; in particular to capture correlation dynamics. However in practice, developing new models faces the "curse of dimensionality" in reference to the - often exponential - increase in the number of model parameters as the number of assets under study grows. Reviews of the multivariate generalised autoregressive conditional heteroskedasticity (GARCH) literature are given by, for example, **?**, **?**, **?** (**?**, Ch. 11) and **?**.

The seed idea in this paper is to undertake a transformation (in particular, a rotation) of the raw returns, which enables us to easily extend the idea of variance targeting (**?**, **?**) to covariance targeting in multivariate models of any dimension. The transformation we propose is not novel, and is related to recent work on the orthogonal GARCH model of **?** and **?**, and its extensions in **?**, **?**, **?** and **?**. The interest in these papers is to find orthogonal or unconditionally uncorrelated components in the raw returns which can then be modelled individually through univariate volatility models.[1] In contrast, we utilise a closely related transformation enabling us to fit flexible multivariate models to the rotated returns using covariance targeting.

We focus on the popular BEKK (**?**, **?**) and Dynamic Conditional Correlations (DCC) (**??**) models, and propose new parameterisations to enrich both models. We focus throughout on diagonal models, to be explained in detail below, and a related parameterisation that offers flexibility in modelling both the volatilities and correlations while economising on the number of parameters. The models we discuss are particularly attractive in terms of estimation and inference, and offers computational advantages compared to existing

[1]The model of **?** differs in that the estimated components are also conditionally uncorrelated. We discuss the relation of our model to orthogonal GARCH models in Section 2.6.

models.

Interest in diagonal models for the DCC process is demonstrated in a number of recent studies, where the objective is to introduce more flexible dynamics while also having parameterisations that are feasible in large dimensions. For p assets, diagonal models in the case of BEKK or DCC, when coupled with covariance targeting, will have a number of dynamic parameters equal to $2p$.[2] In addition to the DCC model with scalar dynamic parameters, **?** also proposed a generalisation with flexible dynamics but it is highly parameterised. Recent studies which focus on DCC with diagonal structures are, for example, **?**, **?**, **?** and **?**.

Within the class of diagonal models, we propose a novel parameterisation that may be attractive in large dimensions. We call it the common persistence (CP) model which imposes common persistence on all elements of the conditional covariance/correlation matrix. This is motivated by the empirical observation that parameter estimates of GARCH(1,1) processes tend to show similar persistence across assets, while exhibiting different levels of smoothness. In addition, the smoothness level seems to change over time; particularly it tends to decline in volatile periods. **?** reports similar findings in his analysis of US financial firms during the recent financial crisis. The common persistence model has only $p+1$ dynamic parameters, and we show that it performs quite favourably in comparison to diagonal models which have $2p$ dynamic parameters.

We show that fitting multivariate volatility models to the rotated returns is essentially the same as fitting models (with different dynamic parameters, in general) to the raw returns; the rotation of the returns simply provides an easier way to do covariance targeting. This equivalence holds since the difference in the likelihood depends on the static parameters needed for the transformation, but is invariant to the type of chosen model. The usefulness of this rotation technique is illustrated using data on the S&P 500 ETF and some DJIA stocks. We analyse bivariate models as well as a moderately large system with 10 DJIA stocks.

The structure of the paper is as follows: Section 2 discusses the model and its properties. Section 3 shows how to estimate the model using a two step estimation strategy, providing a simple multivariate extension of covariance targeting. In Section 4 we apply this model to financial data to illustrate its performance in

[2] We use the term 'dynamic' parameters to denote the parameters of the dynamic equation for the conditional covariance matrix in the BEKK model, and for the conditional correlation matrix in the DCC model. However, covariance targeting also requires the estimation of 'static' parameters which characterise the unconditional second moment of the returns. Estimation is typically undertaken in two stages as discussed later.

comparison to related models. Section 5 draws some conclusions.

2 Modelling Approach

2.1 The Model

First we assume the p-dimensional zero-mean time series

$$r_t, \quad t = 1, ..., T,$$

is ergodic. The unconditional covariance of the returns is given by

$$\text{Var}[r_t] = \overline{H} = P\Lambda P',$$

using the spectral decomposition in the second equality, where P is a matrix of eigenvectors, and the eigenvalue matrix Λ is diagonal with non-negative elements $\lambda_1, \lambda_2, ..., \lambda_p$. Throughout we assume that the eigenvalues in Λ are ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ with $\lambda_p > 0$. It follows that $P^{-1} = P'$ and so $P'P = I$. Hence we can define the symmetric square root of $\overline{H}$

$$\overline{H}^{1/2} = P\Lambda^{1/2}P'.$$

Second, letting $r_t = \overline{H}^{1/2} e_t$ we can define the rotated returns

$$e_t = \overline{H}^{-1/2} r_t = P\Lambda^{-1/2}P' r_t, \quad \text{Var}[e_t] = I.$$

Then we complete the model by specifying the conditional covariance of the rotated returns

$$\text{Var}[e_t | \mathcal{F}_{t-1}] = G_t,$$

where $\text{E}[e_t | \mathcal{F}_{t-1}] = 0$. In order to ease the computational burden, we use a covariance targeting parameterisation (**?**, **?**, in the univariate case of variance targeting) of a BEKK-type model (**?**,**?**) applied to e_t,

$$G_t = \left(I - AA' - BB'\right) + Ae_{t-1}e'_{t-1}A' + BG_{t-1}B', \quad G_0 = I, \tag{1}$$

where we assume

$$\left(I - AA' - BB'\right) \geq 0,$$

in the sense of being positive semidefinite.

Covariance stationarity in (1) follows directly from the analysis of BEKK models by **?** and requires the eigenvalues of $(A \otimes A) + (B \otimes B)$ to be less than one in modulus. Thus unconditionally we can rewrite (1) as

$$\mathrm{E}[G_t] - A\mathrm{E}[G_t]A' - B\mathrm{E}[G_t]B' = I - AA' - BB',$$

where $\mathrm{E}[G_t] = I$ is a solution to this equation implying $\mathrm{E}[e_t e'_t] = I$.

Let $\mathrm{Var}[r_t|\mathcal{F}_{t-1}] = H_t$, fitting the covariance targeting BEKK model to r_t implies

$$H_t = \left(\overline{H} - A\overline{H}A' - B\overline{H}B'\right) + Ar_{t-1}r'_{t-1}A' + BH_{t-1}B', \quad H_0 = \overline{H},$$

which makes estimation challenging in the case of diagonal (when A and B are diagonal) and full (when A and B are unrestricted) BEKK models since it is difficult to impose parameter restrictions to ensure that the target $\left(\overline{H} - A\overline{H}A' - B\overline{H}B'\right)$ is positive semidefinite. Fitting the model to e_t instead, as in (1), circumvents this problem and allows for diagonal and full BEKK models to be easily fitted. In the diagonal case, the parameter restrictions needed for covariance stationarity in (1) also imply a positive semidefinite target.

2.2 Dynamic Properties

The dynamic properties can be studied when the model is vectorised, so we have

$$vec(e_t e'_t) = vec(G_t) + u_t, \quad u_t = vec\left(e_t e'_t - G_t\right),$$

where

$$\begin{aligned} vec(G_t) &= vec\left(I - AA' - BB'\right) + (A \otimes A) vec\left(e_{t-1}e'_{t-1}\right) + (B \otimes B)\, vec\left(G_{t-1}\right) \\ &= vec\left(I - AA' - BB'\right) + \left\{(A \otimes A) + (B \otimes B)\right\} vec\left(G_{t-1}\right) + (A \otimes A) u_{t-1}, \end{aligned}$$

noting that the vector martingale difference property $\mathrm{E}[u_t|\mathcal{F}_{t-1}] = 0$ holds. This implies u_t is a vector weak white noise sequence.

Thus $vec(G_t)$ has a covariance stationary vector autoregression representation while

$$\begin{aligned} vec(e_t e'_t) &= vec\left(I - AA' - BB'\right) + (A \otimes A) vec\left(e_{t-1}e'_{t-1}\right) + (B \otimes B)\, vec\left(G_{t-1}\right) + u_t \\ &= vec\left(I - AA' - BB'\right) + \left\{(A \otimes A) + (B \otimes B)\right\} vec(e_{t-1}e'_{t-1}) \\ &\quad + u_t - (B \otimes B)\, u_{t-1}, \end{aligned}$$

is a covariance stationary vector autoregressive moving average representation.

2.3 Leading Special Cases

2.3.1 Scalar Model

The scalar model specifies $A = \alpha^{1/2} I$ and $B = \beta^{1/2} I$. In this model all elements of G_t have the same dynamic parameters and the dynamic equations are given by

$$g_{ii,t} = (1 - \alpha - \beta) + \alpha e^2_{i,t-1} + \beta g_{ii,t-1}, \quad i = 1, ...p,$$

$$g_{ij,t} = \alpha e_{i,t-1} e_{j,t-1} + \beta g_{ij,t-1}, \quad i, j = 1, ...p,\ i \neq j,$$

where $g_{ij,t}$ denotes the (i, j)-th element of G_t, and we assume $\alpha > 0$ and $\beta \geq 0$. Note that if $\alpha = 0$, β is unidentified and needs to be set equal to zero indicating conditional homoskedasticity in the model, so we rule out this case. To ensure covariance stationarity, we impose $\alpha + \beta < 1$.

2.3.2 Diagonal Model

In this case, A and B are assumed to be diagonal with elements $\alpha_{ii}^{1/2} > 0$ and $\beta_{ii}^{1/2} \geq 0$, respectively. This model implies variance-targeting GARCH(1,1) models for each element of G_t taking the form

$$g_{ii,t} = (1 - \alpha_{ii} - \beta_{ii}) + \alpha_{ii} e_{i,t-1}^2 + \beta_{ii} g_{ii,t-1}, \quad i = 1, ...p,$$

$$g_{ij,t} = \alpha_{ii}^{1/2} \alpha_{jj}^{1/2} e_{i,t-1} e_{j,t-1} + \beta_{ii}^{1/2} \beta_{jj}^{1/2} g_{ij,t-1}, \quad i, j = 1, ...p, \ i \neq j.$$

The cross-equation parameter restrictions between the variance and covariance equations are a feature of BEKK models. Covariance stationarity in this model is determined by the eigenvalues of the diagonal matrix:

$$(A \otimes A) + (B \otimes B) = \begin{pmatrix} \alpha_{11}^{1/2} A + \beta_{11}^{1/2} B & 0 & \cdots & 0 \\ 0 & \alpha_{22}^{1/2} A + \beta_{22}^{1/2} B & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_{pp}^{1/2} A + \beta_{pp}^{1/2} B \end{pmatrix}.$$

Define $\lambda_{ij} = \alpha_{ii}^{1/2} \alpha_{jj}^{1/2} + \beta_{ii}^{1/2} \beta_{jj}^{1/2}$, where λ_{ij} controls the persistence in the (i, j)-th element of G_t.[3] To ensure covariance stationarity, we require that

$$\max \lambda_{ij} < 1, \quad i, j = 1, ...p. \tag{2}$$

In practice, we impose $\lambda_{ii} := \alpha_{ii} + \beta_{ii} < 1$ by reparameterisation, which is a necessary and sufficient condition for (2) to hold; see **?**. This means that in the diagonal BEKK model it suffices to impose covariance stationarity on the conditional variances.

It will be convenient to introduce heterogeneity measures for the smoothness and persistence levels of the elements of G_t. By smoothness we refer to the coefficients β_{ii} for the conditional variances, and $\beta_{ii}^{1/2} \beta_{jj}^{1/2}$ for the conditional covariances, while λ_{ij} is the measure of persistence for the (i, j)-th element of

[3] Recall that the GARCH(1,1) model can be written as $g_{ii,t} = (1 - \alpha_{ii} - \beta_{ii}) + (\alpha_{ii} + \beta_{ii}) g_{ii,t-1} + \alpha_{ii}(e_{i,t-1}^2 - g_{ii,t-1})$, where $e_{i,t-1}^2 - g_{ii,t-1}$ is a martingale difference sequence. Thus the persistence in the conditional variance depends on the autoregression coefficient $(\alpha_{ii} + \beta_{ii})$.

G_t.[4] For ease of interpretation, we do this only for the dynamic parameters of the diagonal elements of G_t (i.e. the conditional variances), noting that the dynamic parameters of the conditional covariance between assets i and j are linked to the dynamic parameters of their conditional variances as shown above. Let $\mu_\alpha = p^{-1}\sum_{i=1}^{p}\alpha_{ii}$ denote the average estimate of α_{ii}, and $\sigma_\alpha = \sqrt{p^{-1}\sum_{i=1}^{p}(\alpha_{ii}-\mu_\alpha)^2}$ be a corresponding measure of heterogeneity. We define similar measures for the smoothness coefficients, β_{ii}, which are μ_β and σ_β, and also for the persistence levels, λ_{ii}, which are denoted by μ_λ and σ_λ. Note that for the scalar model, $\sigma_\alpha = \sigma_\beta = \sigma_\lambda = 0$. These measures are useful for motivating the following model.

2.3.3 Common Persistence (CP) Model

In the diagonal case, $(A \otimes A) + (B \otimes B)$ will be a diagonal matrix with diagonal elements given by $\lambda_{ij} = \alpha_{ii}^{1/2}\alpha_{jj}^{1/2} + \beta_{ii}^{1/2}\beta_{jj}^{1/2}$. The CP model imposes that

$$\lambda_{ij} = \lambda,$$

for all $i, j = 1, ...p$, which gives the dynamic equation

$$G_t = (1-\lambda)I + Ae_{t-1}e_{t-1}'A' + \lambda G_{t-1} - AG_{t-1}A', \tag{3}$$

where A is a diagonal matrix with diagonal elements $0 < \alpha_{ii}^{1/2} < 1$, and $0 < \lambda < 1$ is a scalar parameter satisfying $\lambda > \max \alpha_{ii}$. This model has $p+1$ dynamic parameters as opposed to $2p$ dynamic parameters in the diagonal model. It imposes common persistence on all elements of G_t through a common eigenvalue, λ, for the dynamic equation for G_t. This can be seen from the implied variance-targeting GARCH(1,1) models for each element of G_t given by

$$g_{ii,t} = (1-\lambda) + \alpha_{ii}e_{i,t-1}^2 + (\lambda - \alpha_{ii})g_{ii,t-1}, \quad i = 1, ...p,$$

$$g_{ij,t} = \alpha_{ii}^{1/2}\alpha_{jj}^{1/2}e_{i,t-1}e_{j,t-1} + (\lambda - \alpha_{ii}^{1/2}\alpha_{jj}^{1/2})g_{ij,t-1}, \quad i, j = 1, ...p,\ i \neq j.$$

[4] **?** is interested in similar measures for the conditional variances; however, he defines the smoothness coefficient as $\alpha_{ii}/(\alpha_{ii} + \beta_{ii})$.

The condition for covariance stationarity in this model is simply that $\lambda < 1$, which also implies a positive definite target. The model allows the different elements of G_t to load freely on the lagged variances/covariances and the corresponding shocks allowing them to have different smoothness levels; however it restricts them to have common persistence through λ. In contrast to the diagonal model, here we have $\sigma_\lambda = 0$, while $\sigma_\alpha \neq 0$ which also implies $\sigma_\beta \neq 0$.

This model is motivated by the empirical observation that persistence levels in the conditional variances of asset returns are less heterogeneous compared to their smoothness levels. For instance, **?** studies a large cross section of U.S. financial firms during the 2007-2009 financial crises, and finds the cross-sectional variation in λ_{ii} to be negligible, while smoothness, captured by β_{ii} in our model, tends to decline with the leverage of the company. **?** make a related observation by noting that heterogeneity in α_{ii} is greater than that in β_{ii}, and in one of their models they impose a common smoothing parameter β. We conjecture that imposing a common eigenvalue, λ, is more intuitive since assets with different α_{ii} coefficients are also likely to display varying levels of smoothness through β_{ii}. In addition, the advantage of our specification is that a single parameter, λ, controls both covariance stationarity and positive definiteness of the target regardless of the dimensionality of the system. It also preserves the correlation targeting property which is not the case in the model of **?**.

2.3.4 Orthogonal Parameter Matrices Model

Another interesting case, which we outline here but do not pursue empirically, is when A and B are made up of orthogonal vectors

$$A = (a_1, ..., a_p)', \quad B = (b_1, ..., b_p)'$$

and so

$$\left(AA'\right)_{ij} = a_i' a_j = \alpha_{ij} 1_{[i=j]}, \quad \left(BB'\right)_{ij} = b_i' b_j = \beta_{ij} 1_{[i=j]}, \quad i, j = 1, 2, ..., p,$$

where $1_{[\cdot]}$ is the indicator function. Note that orthogonality of A and B implies that $I - AA' - BB'$ is diagonal. It also implies that $vec(AA') = vec(\Lambda_\alpha)$, where $\Lambda_\alpha = diag\left(\alpha_{11}, ..., \alpha_{pp}\right)$, and similarly for $vec(BB')$.

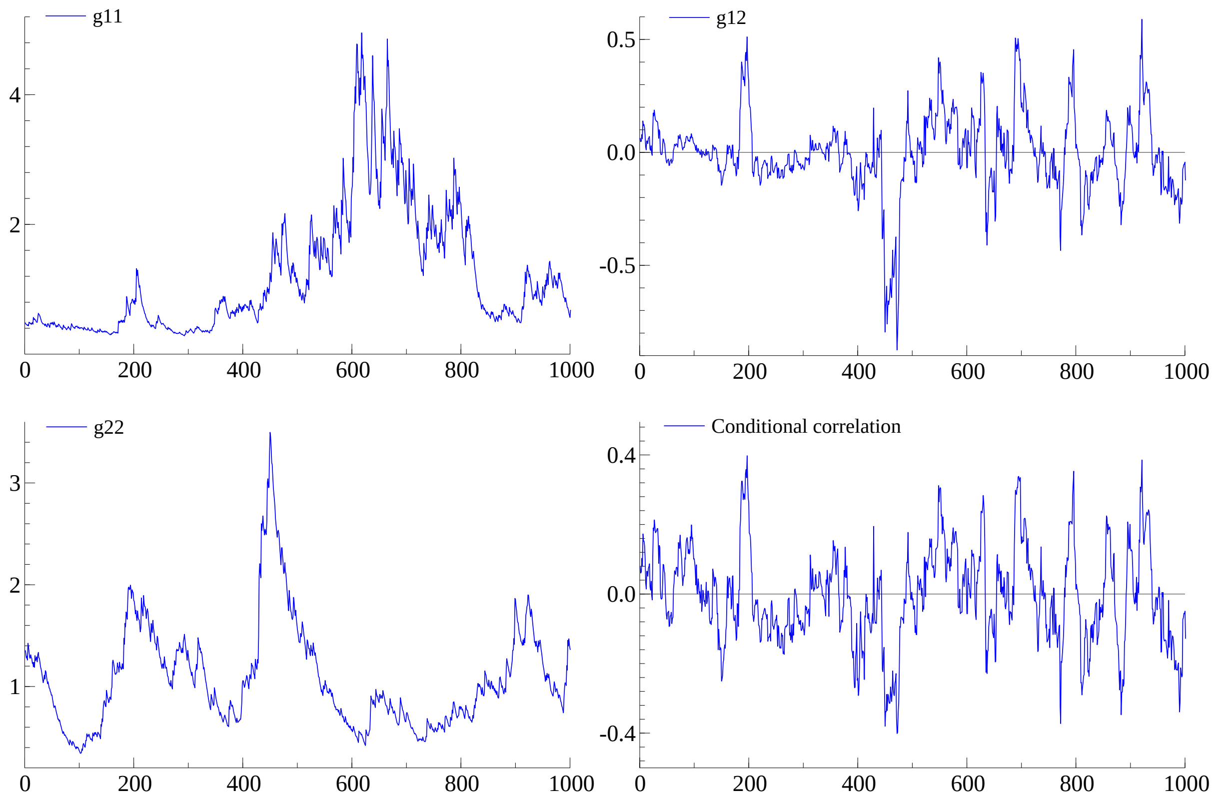

Figure 1: Sample path of G_t for 1,000 simulations.

Example 1. Suppose $p = 2$ and we parameterise the orthogonal case as

$$A = \begin{pmatrix} \alpha_{11}^{1/2} & -c\alpha_{11}^{1/2} \\ c\alpha_{22}^{1/2} & \alpha_{22}^{1/2} \end{pmatrix}, \quad Ae_{t-1} = \begin{pmatrix} \alpha_{11}^{1/2} \\ c\alpha_{22}^{1/2} \end{pmatrix} e_{1,t-1} + \begin{pmatrix} -c\alpha_{11}^{1/2} \\ \alpha_{22}^{1/2} \end{pmatrix} e_{2,t-1},$$

then A is orthogonal and AA' is diagonal with first element $\alpha_{11}\left(1+c^2\right)$, and second element $\alpha_{22}\left(1+c^2\right)$. When $c = 0$ then A is diagonal. In this diagonal case suppose $\alpha_{11}^{1/2} = 0.275$, $\beta_{11}^{1/2} = 0.950$, $\alpha_{22}^{1/2} = 0.200$, $\beta_{22}^{1/2} = 0.980$. Figure 1 shows the sample path of G_t for $1,000$ simulated observations assuming $e_{1,t}$ and $e_{2,t}$ are GARCH(1,1) processes with unconditional variance equal to 1, and persistence levels 0.995 and 0.985, respectively. Top left is $g_{11,t}$, top right is $g_{12,t}$, bottom left is $g_{22,t}$ while bottom right is the implied conditional correlation.

2.4 Implied BEKK Parameterisation

The model in (1) implies that

$$\begin{aligned} \mathrm{Var}[r_t|\mathcal{F}_{t-1}] &= H_t = \overline{H}^{1/2} G_t \overline{H}^{1/2} \\ &= \overline{CC}' + \overline{A} r_{t-1} r_{t-1}' \overline{A}' + \overline{B} H_{t-1} \overline{B}', \end{aligned}$$

where

$$\overline{A} = \overline{H}^{1/2} A \overline{H}^{-1/2}, \quad \overline{B} = \overline{H}^{1/2} B \overline{H}^{-1/2}, \quad \overline{CC}' = \overline{H}^{1/2} \left(I - AA' - BB' \right) \overline{H}^{-1/2}. \tag{4}$$

So this is a particular parameterisation of the **?** BEKK model constructed so it is relatively easy to estimate. It is also clear that this structure does not reproduce an entirely general **?** model; rather it is a constrained version.

Example 2. Suppose A is diagonal, then $\overline{A} = P\Lambda^{1/2}P'AP\Lambda^{-1/2}P'$ which is not symmetric in general. The same applies to $\overline{B}$ when B is diagonal.

This means that when fitting diagonal models to e_t, this implies rather rich dynamics since the implied model for r_t will be a fully parameterised BEKK of the same order. When the asymmetric square root is used (to retrieve the standardised principal components of the data) as in the OGARCH model of **?**, a diagonal model implies $\overline{A} = P\Lambda^{1/2}A\Lambda^{-1/2}P' = PAP'$ which is diagonal, and $\overline{B}$ will also be diagonal. Thus, we prefer the symmetric square root, $P\Lambda^{1/2}P'$, since it will always give a fit that is at least as good as the fit using the asymmetric square root $P\Lambda^{1/2}$.

Example 3. If $A = \alpha^{1/2}I$, then $\overline{A} = \alpha^{1/2}P\Lambda^{1/2}P'P\Lambda^{-1/2}P' = \alpha^{1/2}I$, and the same applies to B. Hence in the scalar case we recover the scalar BEKK model.

It is worth noting that we use BEKK models to model the persistence in G_t, which offers an advantage over the OGARCH and GOGARCH models since the latter models assume that G_t is diagonal; these models are discussed in detail later in Section 2.6. However, we focus on fitting diagonal BEKK models which means the parameters are estimated to fit both the conditional variances and covariances. To the extent that the different elements of G_t have different dynamics, the diagonal BEKK model could potentially lead to a worse fit compared to OGARCH/GOGARCH since the former imposes cross-equation parameter restrictions between the variance and covariance equations. The class of DCC models, which we discuss next, offers more flexibility in this regard and our empirical results indicate its superiority to both BEKK and OGARCH/GOGARCH models.

2.5 DCC Models

2.5.1 Scalar DCC Dynamics

One shortcoming of diagonal and common persistence BEKK models is that the dynamics of $g_{ij,t}$ is linked to the dynamics of $g_{ii,t}$ and $g_{jj,t}$ for all i and j through cross-equation parameter restrictions. This is partly overcome in the DCC model of **?**, which allows for the speed of change in the conditional correlations to be different than that seen for the individual volatilities, and also allows for models to be fit in quite large dimensions. See the discussion in **?**.

DCC models work through first modelling the marginal conditional variances,

$$\mathrm{Var}[r_{i,t}|\mathcal{F}_{t-1}^{r_i}] = \sigma_{i,t}^2, \quad i = 1, 2, ..., p,$$

as univariate GARCH processes. This is an important constraint since in effect it is modelling the conditional variance using its own univariate natural filtration, $\mathcal{F}_{t-1}^{r_i}$. It then computes the standardised potentially correlated innovations

$$v_{i,t} = r_{i,t}/\sigma_{i,t}, \quad i = 1, 2, ..., p.$$

Let $v_t = \left(v_{1,t}, v_{2,t}, ..., v_{p,t}\right)'$ and the unconditional covariance

$$\Pi_C = \mathrm{Var}[v_t],$$

then we model

$$c_{ij,t} = \mathrm{Corr}[v_{i,t}, v_{j,t}|\mathcal{F}_{t-1}], \quad i, j = 1, 2, ..., p.$$

The scalar DCC model decomposes $C_t = [c_{ij,t}]$ as

$$C_t = (Q_t \circ I)^{-\frac{1}{2}} Q_t (Q_t \circ I)^{-\frac{1}{2}},$$

where $\circ$ denotes the Hadamard elementwise product, and Q_t follows a targeted scalar BEKK model

$$Q_t = \left(1 - \alpha - \beta\right) \Pi_C + \alpha v_{t-1} v_{t-1}' + \beta Q_{t-1},$$

where α and β satisfy restrictions similar to the scalar BEKK model; see Section 2.3.1. This ensures that C_t is a genuine correlation matrix.[5] We will call this a scalar DCC model and denote it by S-DCC. The predecessor to the DCC model is the Constant Conditional Correlations (CCC) model of **?** which sets $C_t = C$, where C is the unconditional correlation matrix of v_t.

2.5.2 Flexible DCC Dynamics

The more flexible BEKK-type specifications discussed above suggest similar extensions to the scalar DCC model. Note that a generalisation of the scalar DCC is already mentioned in **?** but it is not pursued empirically. **?** propose more flexible dynamics to the scalar DCC model with asymmetric effects, and they estimate a diagonal DCC model for a group of 34 assets. **?** and **?** fit a restricted diagonal DCC model to 20 assets assuming a sector-specific block structure in A and B such that each matrix has only 3 parameters. **?** introduce a flexible diagonal DCC specification, and apply it to 39 stocks. They overcome the estimation challenge in this high dimension by using a pooled estimator based on **?**.

Based on the standardised returns, v_t, let

$$\Pi_C = \text{Var}[v_t] = P_C \Lambda_C P_C',$$

where P_C contains the eigenvectors and Λ_C has the eigenvalues on the main diagonal. Then we construct

[5] See **?** for a twist on the usual DCC dynamics which has better theoretical properties.

the rotated innovations

$$w_t = P_C \Lambda_C^{-1/2} P_C' v_t.$$

The virtue of this approach is that $\mathrm{Var}[w_t] = I$. Then we model

$$\mathrm{Var}[w_t | \mathcal{F}_{t-1}] = Q_t^*,$$

where

$$Q_t^* = \left(I - AA' - BB'\right) + A w_{t-1} w_{t-1}' A + B Q_{t-1}^* B, \quad Q_0^* = I.$$

As shown for the BEKK parameterisation, Q_t is given by

$$Q_t = P_C \Lambda_C^{1/2} P_C' Q_t^* P_C \Lambda_C^{1/2} P_C'.$$

In the case where $A = \alpha^{1/2} I$ and $B = \beta^{1/2} I$, we reproduce the scalar DCC model. However, the previous sections show we can simply extend this by allowing A and B to be diagonal in the recursion for Q_t^*. This added flexibility may be empirically useful, allowing some aspects of the correlation matrix to move more rapidly than others. We also consider a version similar to the CP model defined in (3).

To conclude, the scalar DCC model is a scalar BEKK model applied to the standardised residuals after fitting univariate GARCH models. The diagonal DCC model, denoted by D-DCC, is simply a diagonal BEKK model applied to the same innovations. The CP model discussed in Section 2.3.3, as a special case of diagonal models with only $p + 1$ dynamic parameters, is somewhat related to one of the proposed models in **?**. The distinction is that **?** impose a common smoothing parameter β on the system, while we impose common persistence through λ. It is important to note that the model of **?** loses the correlation targeting property, while our model preserves this attractive feature.

2.6 Relation to Orthogonal GARCH Models

In this subsection, we take a step back from the different specifications of our model to discuss how it generally relates to some recent propositions known in the literature as orthogonal GARCH models. In writing $r_t = \overline{H}^{1/2} e_t$ and then modelling $e_t = \overline{H}^{-1/2} r_t$, we are effectively analysing the rotated returns. A number of models have focused on linear transformations of the form

$$r_t = Z e_t,$$

where Z is some invertible matrix. Consider the polar decomposition

$$Z = SU, \tag{5}$$

where S is a symmetric positive definite matrix, and U is an orthogonal matrix. Since $\text{Var}[e_t] = I$, we have $\text{Var}[r_t] = ZZ' = S^2$, thus S is the symmetric square root of $\text{Var}[r_t]$ given by $P\Lambda^{1/2}P'$. Therefore part of the matrix Z can be estimated using only unconditional information.

The orthogonal GARCH (OGARCH) model of **?** and **?** assumes $U = P$, hence $Z = P\Lambda^{1/2}$, which is the asymmetric square root of $\text{Var}[r_t]$. In this case e_t is a vector of the standardised principal components of r_t which are unconditionally uncorrelated by construction. **?** assumes that these standardised principal components are also *conditionally* uncorrelated with a diagonal time-varying covariance matrix. This is a misspecification since the standardised principal components will inherit the heteroskedastic properties of the original returns. The generalised OGARCH model (GOGARCH) of **?** proposes $Z = P\Lambda^{1/2}U^*$, where the orthogonal link matrix, U^*, is to be estimated using conditional information. This is sought to avoid identification problems; see **?** for details.

? propose the polar decomposition in (5) and they use conditional information to estimate U under the assumption that some of the estimated components are homoskedastic which leads to dimension reduction.[6] **?** estimate U under the condition that the resulting components, e_t, are also conditionally uncorrelated. Compared to **?**, the models of **?**, **?** and **?** can all be seen as approximations since they assume that the components estimated from their models are conditionally uncorrelated, while in fact they are only unconditionally uncorrelated.

While the model of **?** is conceptually appealing, a set of conditionally uncorrelated components may

[6] **?** focus on a reduced-factor model, while here we focus on the dynamics of the full set of returns.

not exist. Thus, in practice, their model may only give components that are the *least* conditionally correlated in-sample. It is worth noting that estimating the conditionally uncorrelated components in **?** requires solving an $O(p^2)$ optimisation problem which may be infeasible for large dimensions. In addition, they note that as p increases, it becomes more difficult to find factors that are conditionally uncorrelated using their proposed method. **?** adopt a closely related approach to estimate conditionally uncorrelated factors, which departs from earlier work on the GOGARCH model. It is unclear whether their approach guarantees the success of finding conditionally uncorrelated factors in large dimensions.[7]

Our model takes a different stand by directly modelling the conditional covariance matrix of e_t. Here we simply set U in (5) equal to I, which means that e_t is not going to be the only unique set of components satisfying $\text{Var}[e_t] = I$. For instance, we can post-multiply $\overline{H}^{1/2} = P\Lambda^{1/2}P'$ by an arbitrary orthogonal matrix U^*, and still have $\text{Var}[e_t] = \text{E}[e_t e_t'] = \text{E}[U^{*\prime}\overline{H}^{-1/2} r_t r_t' \overline{H}^{-1/2} U^*] = I$. However, uniqueness (or identifiability) of e_t is not crucial since our objective is to simplify estimation and not get unique estimates of e_t. What is important is that for any model for e_t, it is straightforward to derive the implied model for r_t as we discussed in Section 2.4.

For the models we fit, we include the OGARCH model of **?** for comparison. This is equivalent to the following dynamic equation

$$G_t = \left(I - \widetilde{A}\widetilde{A}' - \widetilde{B}\widetilde{B}'\right) + \widetilde{A}\widetilde{A}' \circ \left(e_{t-1}e_{t-1}'\right) + \widetilde{B}\widetilde{B}' \circ G_{t-1}, \tag{6}$$

where $\widetilde{A}$ and $\widetilde{B}$ are diagonal. Note that this equation is for the conditional covariance matrix of the standardised principal components of the returns, i.e. when using the asymmetric square root $\overline{H}^{1/2} = P\Lambda^{1/2}$. We also include results for the GOGARCH model of **?** but modified as proposed by **?**. In this GOGARCH formulation, it is assumed that the transformation matrix, Z, is given by

$$Z = SU(\delta) = P\Lambda^{1/2}P'U(\delta), \tag{7}$$

[7] In all of these studies, the maximum number of assets considered in simulation experiments and for empirical analysis is 12 assets.

where the orthogonal matrix $U(\delta)$ is parameterised by a $p(p-1)/2 \times 1$ vector δ, with j-th element $-180 \leq \delta_j \leq 180$ which is a rotation angle.[8] The dynamics of the resulting $e_t = Z^{-1} r_t$ are modelled as in (6). In models of large dimension, estimating δ is generally challenging, thus we only include the GOGARCH model for comparison in our empirical analysis in the bivariate case. Note that our model imposes $U(\delta) = I$, or equivalently $\delta = 0$.

To summarise, the key feature of OGARCH and GOGARCH models is that conditionally the factors are assumed to be uncorrelated. This is not true of (1), which assumes they follow a BEKK-type model. The models are not the same even in the scalar BEKK case, hence these models are non-nested. In (1) when A and B are diagonal, the diagonal elements of G_t follow similar dynamics to the OGARCH/GOGARCH model as in (6). The models differ by the non-diagonal elements of G_t which are always assumed to be zero in the OGARCH/GOGARCH structure. This means that the marginal likelihoods for the univariate series $e_{i,t}$, $i = 1, ..., p$, are the same (holding the parameters equal across the models) but their dependence structure will be different.

2.7 A Time-Varying-Weight Strict Factor Model Representation

Our model can also be interpreted as a time-varying-weight strict factor model. The model implies

$$\begin{aligned} \text{Var}[r_t|\mathcal{F}_{t-1}] &= H_t = \overline{H}^{1/2} G_t \overline{H}^{1/2} \\ &= P\Lambda^{1/2} P' G_t P \Lambda^{1/2} P'. \end{aligned}$$

Suppose we take the spectral decomposition of G_t at each point in time such that $G_t = P_t^G \Lambda_t^G (P_t^G)'$, where P_t^G contains the eigenvectors of G_t and the diagonal matrix Λ_t^G has the eigenvalues of G_t along its main diagonal. Then we can write

$$H_t = P\Lambda^{1/2} P' P_t^G \Lambda_t^G (P_t^G)' P \Lambda^{1/2} P'$$

[8]Note that any 2×2 orthogonal matrix can be written as a rotation matrix taking the form $U(\delta) = \begin{pmatrix} \cos\delta & -\sin\delta \\ \sin\delta & \cos\delta \end{pmatrix}$ where $-180 \leq \delta \leq 180$, which is scalar in this example, is a rotation angle. A positive δ indicates counterclockwise rotation. For $p > 2$, $U(\delta)$ can be represented as the product of $p(p-1)/2$ rotation matrices each paramterised with a distinct rotation angle; see **?** for details.

$$= z_t \Lambda_t^G z_t',$$

where z_t is a time-varying weight matrix. This representation is reminiscent of strict factor models where the factors are not correlated, their conditional variances are given by the diagonal elements of the time-varying Λ_t^G, and there is no approximation error covariance since the number of factors is equal to the number of assets.

The term strict factor model is usually used to characterise a model where the idiosyncratic components of asset returns are uncorrelated as in **?**, for example. Here we adapt it to describe a model where the factors are uncorrelated both conditionally and unconditionally, and the factor loadings, z_t, are time-varying.

Note that orthogonal GARCH models assume that G_t is diagonal, and in this case $\Lambda_t^G = G_t$ while $P_t^G = I$. Thus orthogonal GARCH models impose a fixed weight matrix $z_t = z$. This representation provides an additional intuition behind our model, and explains why capturing the covariance dynamics of $e_{i,t}$, $i = 1, ..., p$, is important. Since we also consider DCC-type parameterisations, this analogy can be extended to the factor DCC model of **?** which, if reparameterised as above, becomes a time-varying-weight strict factor model.

3 Inference

3.1 Parameter Vector

We will focus on the two part model, where the first part is

$$\mathrm{E}[r_t] = 0, \quad \mathrm{Var}[r_t] = \overline{H} = P\Lambda P', \quad t = 1, 2, ..., T,$$

and the second is

$$e_t = P\Lambda^{-1/2}P'r_t, \quad \mathrm{E}[e_t|\mathcal{F}_{t-1}] = 0, \quad \mathrm{Var}[e_t|\mathcal{F}_{t-1}] = G_t,$$

and

$$G_t = \left(I - AA' - BB'\right) + Ae_{t-1}e_{t-1}'A' + BG_{t-1}B', \quad G_0 = I.$$

Let θ_A and θ_B denote the parameters indexing A and B. The parameters in the model are

$$\theta = \left(vech(\overline{H})', \theta_A', \theta_B'\right)' = \left(\theta_{\overline{H}}', \theta_*'\right)'.$$

We call θ_* the 'dynamic' parameters and $\theta_{\overline{H}}$ the 'static' parameters. The true values of these parameters are denoted by $\theta_{0,*}$ and $\theta_{0,\overline{H}}$, respectively, while $\theta_0 = \left(\theta_{0,\overline{H}}', \theta_{0,*}'\right)'$. Typically the dimension of $\theta_{\overline{H}}$ is large and potentially massive if p is large since it has $O(p^2)$ elements. The dimension of θ_* is often small with only $O(p)$ parameters in the specifications we consider.

In the diagonal case, let $\theta_{*,i}$ denote the dynamic parameters which index the dynamics of the i-th series $e_{i,t}$. Thus $\theta_{*,i} = \left(\alpha_{ii}, \beta_{ii}\right)$, and $\theta_* = \left(\theta_{*,1}', \theta_{*,2}', ..., \theta_{*,p}'\right)'$, recalling that p is the number of assets. This notation will be useful later when discussing the numerical optimisation algorithm we use for diagonal models.

3.2 Two Step Estimation

The structure of the model allows for a two-step estimation strategy to estimate θ. This approach, which dramatically eases the computational burden, was advocated in the univariate case by **?** and has been used for the scalar BEKK and DCC models in many studies.

In the first step we focus solely on the static parameters $\theta_{\overline{H}} = vech(\overline{H})$. By construction $\overline{H} = \text{Var}[r_t]$, thus we use the method of moments estimator

$$\widehat{\overline{H}} = \frac{1}{T}\sum_{t=1}^{T} r_t r_t',$$

implying $\widehat{\theta_{\overline{H}}}$. This estimate is then decomposed into $\widehat{P}$ and $\widehat{\Lambda}$. Then we construct the time series of rotated returns

$$e_t = \widehat{P}\widehat{\Lambda}^{-1/2}\widehat{P}' r_t, \quad t = 1, 2, ..., T.$$

The second stage estimation is based on the quasi-likelihood

$$\log L(\theta_*, \widehat{\theta}_{\overline{H}}) = \sum_{t=1}^{T} \log L_t(\theta_*, \widehat{\theta}_{\overline{H}}) = const - \frac{1}{2}\sum_{t=1}^{T}\log|G_t| - \frac{1}{2}\sum_{t=1}^{T} e_t' G_t^{-1} e_t, \tag{8}$$

where

$$G_t = \left(I - AA' - BB'\right) + Ae_{t-1}e_{t-1}'A' + BG_{t-1}B', \quad G_0 = I. \tag{9}$$

This is optimised solely over θ_*, keeping $\widehat{\theta}_{\overline{H}}$ fixed, which delivers $\widehat{\theta}_*$. If the dimensionality of the system, p, becomes very large then it may be worth switching over to use a composite likelihood (**?**, **?**, and **?**, **?**) or the McGyver estimation method (**?**, **?**), but we will not discuss that here.

When estimating the OGARCH model, we use $e_t = \widehat{\Lambda}^{-1/2}\widehat{P}' r_t$ in (8) while the dynamic equation (9) is replaced with (6). For GOGARCH we use $e_t = \widehat{P}\widehat{\Lambda}^{-1/2}\widehat{P}' r_t$ in the following quasi-likelihood

$$\log L(\theta_*, \widehat{\theta}_{\overline{H}}) = \sum_{t=1}^{T} \log L_t(\theta_*, \widehat{\theta}_{\overline{H}}) = const - \frac{1}{2}\sum_{t=1}^{T}\log|G_t| - \frac{1}{2}\sum_{t=1}^{T} e_t' U(\delta) G_t^{-1} U(\delta)' e_t, \tag{10}$$

and the dynamic equation for G_t is also given by (6). In this case, the additional $p(p-1)/2$ in δ are contained in θ_*.[9]

In terms of asymptotic theory, for fixed p and $T \to \infty$, this is simply a two step moment estimator, e.g. **?** and **?**, where the moment conditions are given by the vector

$$\begin{aligned} m(\theta_*, \theta_{\overline{H}}) &= \sum_{t=1}^{T} m_t(\theta_*, \theta_{\overline{H}}), \quad m_t(\theta_*, \theta_{\overline{H}}) = \begin{pmatrix} \theta_{\overline{H}} - vech\left(r_t r_t'\right) \\ \\ \frac{\partial \log L_t(\theta_*, \theta_{\overline{H}})}{\partial \theta_*} \end{pmatrix}, \\ m(\widehat{\theta}_*, \widehat{\theta}_{\overline{H}}) &= 0, \end{aligned}$$

and

$$\mathrm{E}\left\{ m(\theta_*, \theta_{\overline{H}})\big|_{\theta_*=\theta_{0,*};\theta_{\overline{H}}=\theta_{0,\overline{H}}} \right\} = 0,$$

[9]As noted earlier the estimation of δ is challenging when p is large. Thus we only estimate the GOGARCH model in the bivariate case in our empirical analysis.

at the true values. The key feature here is that the first step does not involve θ_*, which simplifies the estimation of the dynamic parameters in the second step.

The asymptotic distribution of this two step estimator has been worked over by many authors in the context of scalar BEKK models and the DCC model, so we will not discuss it in detail here. Under standard regularity conditions, as $T \to \infty$ we have

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}(\mathcal{I}^{-1})')$$

where $\widehat{\theta} = \left(\widehat{\theta}_*', \widehat{\theta}_{\overline{H}}'\right)'$,

$$\mathcal{J} = \text{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} m_t(\theta_*, \theta_{\overline{H}})\right], \quad \mathcal{I} = \text{E}\left[\frac{\partial m_t(\theta_*, \theta_{\overline{H}})}{\partial \theta'}\right],$$

and we use a HAC estimator, e.g. **?**, to estimate $\mathcal{J}$.

3.3 Numerical Optimisation

General numerical optimisation routines can be used to locate $\widehat{\theta}_*$. An alternative, which we have used systematically in the estimation of diagonal models, is to employ a zig-zag algorithm based upon the structure of θ_*. We optimise

$$\log L(\theta_{*i}, \theta_{*\backslash i}, \widehat{\theta}_{\overline{H}}),$$

with respect to θ_{*i}, holding all other elements of θ_*, written as $\theta_{*\backslash i}$, at the previously best values. We then cycle over i, repeating the optimisation each time. The advantage of this is that each individual optimisation is only 2-dimensional, and we have found this method to be reliable. The theory for this estimator is discussed in **?**, while inference is standard as outlined in Section 3.2.

3.4 Model Comparison

We will use a quasi-likelihood criterion for r_t to compare the fit of the different models, which means we will focus on the 1-step prediction ability of the models using a Kullback-Leibler distance. Note that given the likelihood for e_t, it is straightforward to compute the likelihood for r_t since the Jacobian of the transformation is $\frac{\partial r_t}{\partial e_t'} = P\Lambda^{1/2}P'$, and its determinant is $\left|P\Lambda^{1/2}P'\right| = \left|\Lambda^{1/2}\right|$, where the second equality follows from P being orthogonal; see **?** (**?**, pp. 48). Thus for a time series of length T, we have that $\log L_r = \log L_e - \frac{T}{2}\log|\Lambda|$, where $\log L_r$ and $\log L_e$ denote the log-likelihoods for r_t and e_t, respectively. This also implies that comparisons based on models for e_t is equivalent to comparisons based on equivalent models for r_t. This is because the difference in the likelihood is independent of the dynamic parameters, and only depends on the static parameters, Λ, which are common to all the models we consider.

Let $\log L_{a,t}$ denote the t-th observation log-likelihood for r_t based on model a. To compare two models, a and b, we look at the average log-likelihood difference

$$l_{a,b} = \frac{1}{T}\sum_{t=1}^{T} l_{a,b,t}, \quad l_{a,b,t} = \log L_{a,t} - \log L_{b,t}. \tag{11}$$

We then test if $l_{a,b}$ is statistically significantly different than zero by computing a HAC estimator of the variance of $l_{a,b}$. This predictive ability test was first introduced by **?**. Using a quasi-likelihood criterion is valid for non-nested and misspecified models; see **?** and **?** for in-sample model comparison, and **?** for out-of-sample model selection. For comparisons, we choose the diagonal model within each class (BEKK, DCC, OGARCH and GOGARCH) since it is the most flexible specification, and then test for equal predictive ability. We will use either OGARCH or GOGARCH in a comparison, but not both since the GOGARCH nests OGARCH and thus this test would not be appropriate.[10]

3.5 Copula and Marginal Likelihoods

It is also useful to consider the marginal log-likelihood for the i-th series

$$\log L_i = \sum_{t=1}^{T} \log f(r_{i,t}|\mathcal{F}_{t-1}),$$

[10]If interest is in testing nested models, the approach of **?** can be adopted by using rolling-window estimation. This allows for pairwise comparisons of the predictive ability of all the four classes of models as well as the different variants under each class.

where we have conditioned on the entire filtration, not just the natural filtration for the i-th series. The implied copula likelihood is then given by

$$\log L - \sum_{i=1}^{p} \log L_i.$$

Under the assumption of conditional normality, the copula parameter is the conditional correlation matrix of the returns. For the copula-margins decomposition in the CCC and DCC models, see, respectively, equation (6) in **?** and equation (26) in **?**.

4 Empirical Analysis

4.1 Data

We use close-to-close daily returns data on Spyder (SPY), an S&P 500 exchange traded fund, and some of the most liquid stocks in the Dow Jones Industrial Average (DJIA) index. These are: Alcoa (AA), American Express (AXP), Bank of America (BAC), Coca Cola (KO), Du Pont (DD), General Electric (GE), International Business Machines (IBM), JP Morgan (JPM), Microsoft (MSFT), and Exxon Mobil (XOM). The sample period is 1/2/2001 to 31/12/2009 and the source of the data is Yahoo!Finance, which is accessible online. We use close prices adjusted for dividends and splits.

Our primary empirical example in Section 4.3 focuses on the pair XOM-AA, which we use to present the models' main features. In Section 4.4, we analyse stock-index dynamics by studying the pair SPY-XOM. This sheds light on the conditional correlation of a firm's stock with the overall market index, and the latter part of our sample includes the recent financial crisis. This application relates to the recent work of **?** and **?** where they focus on modelling systemic risk measures using conditional correlations and conditional betas, respectively. See also **?** for a multivariate volatility model for the same group of assets which utilises high frequency data. In Section 4.5 we estimate the models using all 10 stocks from the DJIA index.

4.2 Considered Models

4.2.1 BEKK Class

We work with the rotated returns $e_t = \widehat{P}\widehat{\Lambda}^{-1/2}\widehat{P}' r_t$, which are unconditionally uncorrelated in-sample and each has unconditional variance equal to 1. They display, of course, volatility clustering. Then we fit the covariance targeting BEKK model

$$\text{Var}[e_t|\mathcal{F}_{t-1}] = G_t = \left(I - AA' - BB'\right) + Ae_{t-1}e_{t-1}'A' + BG_{t-1}B', \quad G_0 = I.$$

The dynamics are estimated using a Gaussian quasi-likelihood. We fit the following models:

- Scalar BEKK (S-BEKK). $A = \alpha^{1/2} I$ and $B = \beta^{1/2} I$.
- Diagonal BEKK (D-BEKK). $A = diag(\alpha_{11}^{1/2}, ..., \alpha_{pp}^{1/2})$ and $B = diag(\beta_{11}^{1/2}, ..., \beta_{pp}^{1/2})$.
- Diagonal BEKK with common persistence (D-BEKK-CP). $A = diag(\alpha_{11}^{1/2}, ..., \alpha_{pp}^{1/2})$ and λ is the common persistence parameter.

For comparison, we also report results for these three specifications when applied to OGARCH-type and GOGARCH-type models, where in the latter models it is assumed that G_t is diagonal.[11] The diagonal OGARCH and GOGARCH models (with unconstrained $\alpha_{ii}^{1/2}$ and $\beta_{ii}^{1/2}$) correspond to the models of **?** and **?**, respectively, while the other specifications are novel in this context.

4.2.2 DCC Class

We first fit variance targeting univariate GARCH(1,1) models to the returns, which produces a sequence of standardised vector innovations v_t. Then we model $c_{ij,t} = \text{Corr}[v_{i,t}, v_{j,t}|\mathcal{F}_{t-1}]$. The conditional correlation matrix $C_t = [c_{ij,t}]$ is decomposed as

$$C_t = (Q_t \circ I)^{-\frac{1}{2}} Q_t (Q_t \circ I)^{-\frac{1}{2}}.$$

[11] Note that the OGARCH model is for $e_t = \widehat{\Lambda}^{-1/2}\widehat{P}' r_t$, while the GOGARCH model is for $e_t = \widehat{P}\widehat{\Lambda}^{-1/2}\widehat{P}' r_t$ and the latter's likelihood is given by (10).

We first rotate v_t to generate $w_t = P_C \Lambda_C^{-1/2} P_C' v_t$, then we model $\text{Var}[w_t|\mathcal{F}_{t-1}] = Q_t^*$ and then take $Q_t = P_C \Lambda_C^{1/2} P_C' Q_t^* P_C \Lambda_C^{1/2} P_C'$. The dynamic equation for Q_t^* is

$$Q_t^* = \left(I - AA' - BB'\right) + A w_{t-1} w_{t-1}' A + B Q_{t-1}^* B, \quad Q_0^* = I,$$

which is estimated using a Gaussian quasi-likelihood. We estimate the following models:

- Constant conditional correlations (CCC). $A = B = 0$.
- Scalar DCC (S-DCC). $A = \alpha^{1/2} I$ and $B = \beta^{1/2} I$.
- Diagonal DCC (D-DCC). $A = diag(\alpha_{11}^{1/2}, ..., \alpha_{pp}^{1/2})$ and $B = diag(\beta_{11}^{1/2}, ..., \beta_{pp}^{1/2})$.
- Diagonal DCC with common persistence (D-DCC-CP). $A = diag(\alpha_{11}^{1/2}, ..., \alpha_{pp}^{1/2})$ and λ is the common persistence parameter.

4.3 Analyzing the Pair XOM-AA

4.3.1 BEKK, OGARCH and GOGARCH Models

We will start out with a detailed bivariate example: the daily returns of Exxon Mobil (XOM) and Alcoa (AA). Figure 2 provides summary of the series. The daily returns are in the upper panel while the rotated returns are in the lower panel. The unconditional covariance matrix of the returns is given in Table **??**. The first eigenvector looks like a market factor, while the second is a long/short portfolio.

	Returns				GARCH(1,1) innovations			
	Covariance		Eigenvectors		Covariance		Eigenvectors	
XOM	3.069	2.918	0.394	-0.919	0.981	0.470	0.706	-0.708
AA	2.918	8.633	0.919	0.394	0.470	0.983	0.708	0.706
Eigenvalues	–	–	9.882	1.820	–	–	1.452	0.512

Table 1: Left-hand side is the unconditional covariance of returns, together with their eigenvalues and (normalised) eigenvectors. On the right-hand side is the unconditional covariance of the innovations from univariate variance targeting GARCH(1,1) models.

The parameter estimates of the BEKK, OGARCH and GOGARCH models are given in Table **??** together with the associated log-likelihood values for the (unrotated) returns evaluated at $(\widehat{\theta}_*, \widehat{\theta}_{\overline{H}})$. The joint log-likelihood is decomposed to indicate the performance in terms of the margins and the copula. In the BEKK

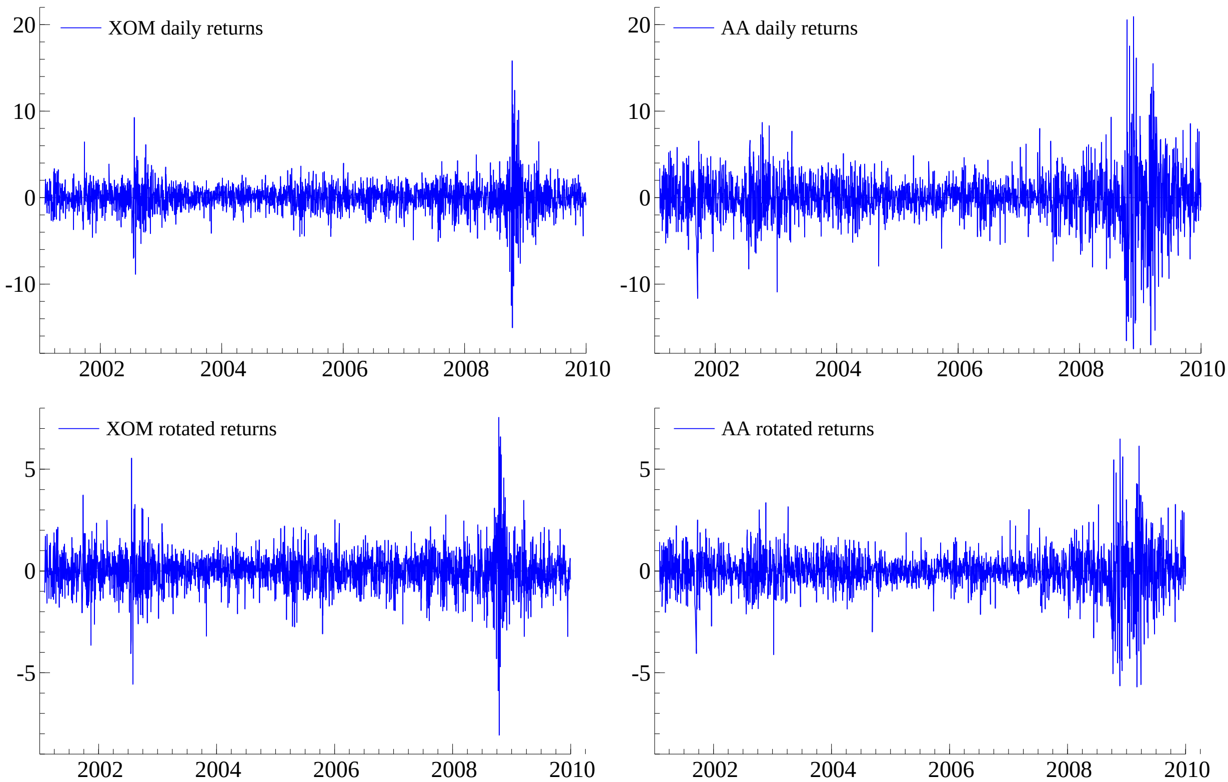

Figure 2: XOM and AA series: Top panel plots the daily returns (r_t). Bottom panel plots the rotated returns (e_t).

class, the D-BEKK model provides a moderate improvement in fit compared to S-BEKK. This is due to the diagonal parameters freely fitting each conditional variance. The effects are quite considerable since α_1 and α_2 are an order of magnitude different than the S-BEKK's α, so that XOM's conditional variance dynamics are much more responsive to its own shock, while the estimates for the conditional variance of AA are smoother. Of course, these estimates also fit the conditional covariance dynamics given the cross-equation parameter restrictions of the diagonal BEKK model.

The parameters of the implied BEKK model for r_t, given by (4), are

$$\overline{A} = \overline{H}^{1/2} A \overline{H}^{-1/2} = \begin{pmatrix} 0.275 & -0.019 \\ 0.034 & 0.182 \end{pmatrix}, \quad \overline{B} = \overline{H}^{1/2} B \overline{H}^{-1/2} = \begin{pmatrix} 0.951 & 0.007 \\ -0.012 & 0.983 \end{pmatrix},$$

indicating that a diagonal model for the rotated returns implies a full BEKK model for the unrotated returns.

	BEKK			OGARCH			GOGARCH		
	Scalar	Diagonal	CP	Scalar	Diagonal	CP	Scalar	Diagonal	CP
α	0.050	–	–	0.066	–	–	0.063	–	–
β	0.943	–	–	0.924	–	–	0.927	–	–
α_{11}	–	0.072	0.055	–	0.060	0.076	–	0.088	0.067
α_{22}	–	0.036	0.044	–	0.082	0.059	–	0.041	0.055
β_{11}	–	0.909	–	–	0.934	–	–	0.884	–
β_{22}	–	0.960	–	–	0.887	–	–	0.955	–
λ	–	–	0.993	–	–	0.989	–	–	0.991
δ	–	–	–	–	–	–	0.015	-0.008	0.030
LL decomposition									
Margin (XOM)	-4,034	-4,030	-4,032	-4,028	-4,031	-4,027	-4,029	-4,030	-4,028
Margin (AA)	-5,098	-5,098	-5,099	-5,099	-5,126	-5,102	-5,100	-5,097	-5,099
Copula	284	288	284	233	270	235	258	266	258
Total LL	-8,848	-8,840	-8,847	-8,894	-8,887	-8,893	-8,870	-8,861	-8,869

Table 2: Dataset: XOM and AA daily returns 1/2/2001-31/12/2009. Top Panel: parameter estimates of the scalar, diagonal, and common persistence (CP) parameterisations for the BEKK, OGARCH and GOGARCH models. α and β are the parameters of the scalar models, while $(\alpha_{ii}, \beta_{ii})$, $i = 1, 2$, are those of the diagonal models. For CP, λ (the common persistence parameter) and α_{ii} for each asset are reported. δ is the rotation angle in the bivariate GOGARCH model. All parameters, except for δ, are statistically significant at 5 percent. Bottom panel: Log-likelihood decomposition at the estimated parameter values.

Recall that this follows from specifying $\overline{H}^{1/2}$ as the symmetric square root using the spectral decomposition. The D-BEKK-CP model estimates imply roughly the same level of persistence in the elements of G_t as the S-BEKK and D-BEKK models. The picture for OGARCH and GOGARCH is rather similar but indicating a slightly lower level of persistence.

Interestingly, the GOGARCH model's estimated rotation angle is very close to zero and statistically insignificant. This implies that $U(\delta) \approx I$, making the e_t series from the GOGARCH model very close to those from the BEKK model; see (7). The primary difference between the two models is that GOGARCH assumes that $g_{12,t}$ is zero, which is reflected in BEKK's superior copula fit.

The BEKK models provide an important increase in the likelihood compared to OGARCH and GOGARCH. The increase in the log-likelihood in BEKK models is primarily due to an increase in the copula fit, implying that capturing the conditional correlations in the rotated returns (which is not the case in OGARCH and GOGARCH) does improve the modelling of the conditional correlations of the unrotated returns. There is a small loss in fit in the first margin (XOM) when using the BEKK model, however this is more than compensated through capturing the conditional correlation dynamics with BEKK models providing an overall gain in fit.

4.3.2 DCC Models

Table **??** gives estimates of the CCC and DCC models. When estimating the variance targeting GARCH(1,1) models for the margins, we first standardise the returns of XOM and AA by their respective unconditional variances, fit variance targeting GARCH(1,1) models for these standardised returns and report the log-likelihood for the original returns as the marginal log-likelihood. The estimates suggest different dynamics for the two series, which can already be inferred from the improvement offered by the diagonal models in Table **??**. Not surprisingly, the fit for the margins in this case is better than all the BEKK, OGARCH and GOGARCH models. For CCC the unconditional correlation of the standardised returns is 0.480. We use the unconditional correlation to build the time-varying covariance matrix, the dynamics of which are driven only by the conditional variances in this model.

		DCC		
	CCC	Scalar	Diagonal	CP
Variance parameters				
Margin (XOM)		(0.084, 0.901)		
Margin (AA)		(0.048, 0.948)		
Correlation parameters				
CCC	0.480	–	–	–
α	–	0.015	–	–
β	–	0.977	–	–
α_{11}	–	–	0.006	0.005
α_{22}	–	–	0.037	0.054
β_{11}	–	–	0.993	–
β_{22}	–	–	0.960	–
λ	–	–	–	0.992
LL decomposition				
Margin (XOM)	-4,026	-4,026	-4,026	-4,026
Margin (AA)	-5,096	-5,096	-5,096	-5,096
Copula	293	306	307	307
Total LL	-8,829	-8,816	-8,815	-8,815

Table 3: Dataset: XOM and AA daily returns 1/2/2001-31/12/2009. Parameter estimates of the constant conditional correlations (CCC), and scalar, diagonal and common persistence (CP) parameterisations for the DCC model. Top panel: estimates of the variance targeting GARCH(1,1) models for the margins. Middle panel: estimates of the correlation parameters: α and β are the parameters of S-DCC, while $(\alpha_{ii}, \beta_{ii})$, $i = 1, 2$, are those of D-DCC. For CP, λ (the common persistence parameter) and α_{ii} for each asset are reported. All parameters are statistically significant at the 5 percent level of significance. Bottom panel: Log-likelihood decomposition at the estimated parameter values.

The estimates for the DCC dynamics suggest only a marginal improvement by the D-DCC and D-DCC-CP over S-DCC. With the margins fit freely, there seems to be no additional improvement from further enriching the DCC dynamics in this case. This is in contrast to the BEKK model results, but it is perhaps unsurprising since there is a single conditional correlation to model in this case. As we show later, in higher dimensions the gains from the further flexibility of the D-DCC and D-DCC-CP models can be substantial. Overall the estimates suggest that the conditional correlation matrix is quite persistent. The log-likelihood decomposition results indicate a rather significant improvement in the overall fit compared to the BEKK, OGARCH and GOGARCH models, especially in comparison to OGARCH.

Figure 3 plots the conditional correlations from the diagonal models which provided the best fit in each model class. The D-DCC conditional correlation is the most persistent and lies within a tighter range. It appears to be generally lower than the conditional correlation from the D-BEKK and D-OGARCH model, with the exception of the year 2005 where D-OGARCH conditional correlation was noticeably lower. This

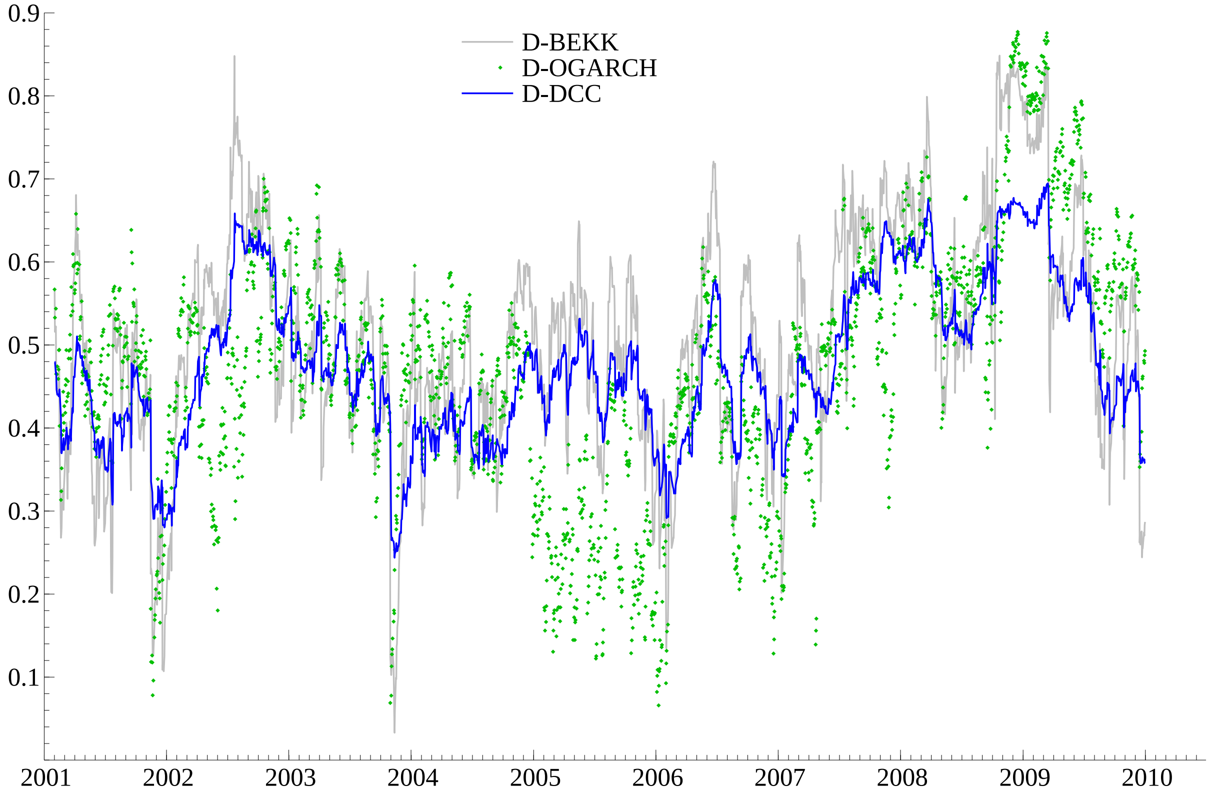

Figure 3: Conditional correlations from the diagonal BEKK, OGARCH and DCC models.

observation is perhaps most evident during the latter part of the financial crisis, roughly starting 2009, with the difference in the implied correlation level being rather significant at times during this period.

We apply the 1-step predictive ability test outlined in Section 3.4 to the D-BEKK, D-OGARCH and D-DCC models which are the most flexible in each class. Comparing D-BEKK to D-OGARCH gives a t-statistic of 2.81 which is statistically significant at 1 percent, indicating that D-BEKK provides superior 1-step forecasts. Comparing D-DCC to D-BEKK and D-OGARCH gives t-statistics equal to 2.24 and 3.74, respectively, indicating that D-DCC outperforms both models out of sample. These results are, of course, in line with the substantial in-sample gains shown by the DCC models.

4.4 Index-Stock Dynamics: SPY-XOM

The results for SPY-XOM are reported in Table **??**. Moving from S-BEKK to D-BEKK leads to a modest improvement in fit for the first margin and the copula. This is also the case in OGARCH with gains only in

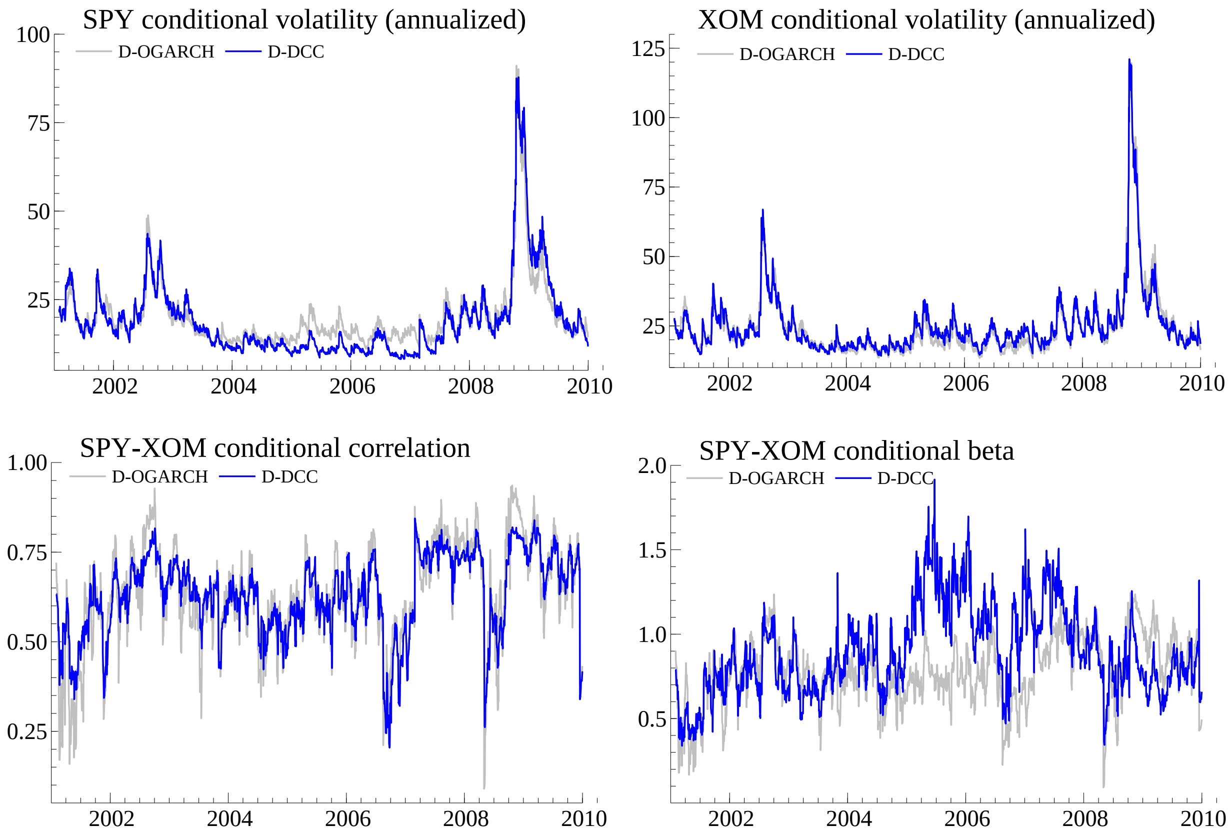

Figure 4: SPY-XOM conditional variances, correlation and beta from the diagonal OGARCH and DCC models.

the first margin. GOGARCH provides considerable gain compared to OGARCH, particularly in the copula fit, with a statistically significant estimate of the rotation angle at about -130 degrees. Both DCC and BEKK models improve significantly over OGARCH and GOGARCH, with DCC providing some gain over BEKK in both margins and the copula.

In terms of predictive ability, the D-BEKK model provides superior 1-step forecasts compared to OGARCH models with a t-statistic of 4.48. The D-DCC also significantly improves over D-OGARCH with a t-statistic of 4.77; however, its improvement over D-BEKK is statistically insignificant with a t-statistic of 1.12. Again this mirrors the in-sample results of the three models.

Figure 4 shows the conditional volatilities, correlation and beta from D-OGARCH and D-DCC for SPY-XOM. The conditional variances from the two models seem quite similar, except for the SPY conditional volatility during 2005-2007 where the difference is mainly one of scale. The path of the conditional correlations is also somewhat similar although the D-OGARCH model attains more spikes. The interesting difference in this figure is the rather different profile for the conditional beta. From 2005 to mid 2007, the D-DCC model implies a conditional beta that is consistently larger and typically greater than 1, and it seems to have moderated gradually during the financial crisis.

4.5 Ten Dimensional Example

We now analyse all 10 stocks from the DJIA index. The first two eigenvectors, corresponding to the two largest eigenvalues of the unconditional covariance matrix of the returns, are reported in Table **??**. The first eigenvector looks roughly like a market factor and the second is a market portfolio that is short (long) in financial stocks (BAC, JPM and AXP) and long (short) in the other stocks.[12] The two largest eigenvalues are 35.93 and 6.85, and they account for 73 percent of the total variation in the returns, where total variation is measured by the trace of $\overline{H}$.

Table **??** shows the estimated parameters for the scalar, diagonal and common persistence models. The latter are an interesting alternative in moderately large dimensions since they have only $p+1$ dynamic parameters compared to $2p$ parameters in the diagonal models. Moving from the scalar to the diagonal models seems to pay off with a considerable improvement in overall fit in D-BEKK, and less so for D-OGARCH. The BEKK models provide a significant overall gain in the log-likelihood over OGARCH all due to improving the copula fit. Note that the BEKK loses in the margins to OGARCH as the BEKK parameters provide a fit to both the variance and covariance elements of G_t.

Of course, the DCC models provide the best fit since the margins are freely estimated. The overall gain compared to BEKK and OGARCH is quite impressive, and the DCC gains are uniform across all margins and the copula. Unlike BEKK and OGARCH cases, moving from S-DCC to D-DCC does not improve the copula fit massively. In this moderately large dimension, the favourable performance of the CP model is evident, particularly in the BEKK and OGARCH cases. In both cases, the diagonal specifications significantly improve the overall fit (mostly due to the copula contribution) and when fitting the CP model the deterioration in fit is rather slight. To a lesser extent, this is also the case in the DCC models.

Given that both the scalar and CP specifications are nested in the diagonal model, we can use a likeli-

[12]Note that when the eigenvalues are distinct, the normalised eigenvectors are unique up to sign.

hood ratio (LR) test. The scalar model imposes $2p$ restrictions on the diagonal model, and according to the LR test, the reduction in fit is statistically significant at 5 percent in all three cases. The CP model imposes $p(p+1)/2$ restrictions on the diagonal model and according to the LR test, the loss in fit when moving from D-BEKK to D-BEKK-CP is statistically significant at 5 percent, while this is not the case in the OGARCH and DCC models.

This is an interesting result since the number of dynamic parameters in the CP model is $p+1$ compared to $2p$ dynamic parameters in the diagonal model. This could be due to the differences in the heterogeneity in the persistence and smoothness levels among the parameters of the diagonal models. For instance, in D-BEKK the heterogeneity in the parameters is given by $\sigma_\alpha = 0.014$ and $\sigma_\beta = 0.023$, while the corresponding measures in D-DCC are $\sigma_\alpha = 0.004$ and $\sigma_\beta = 0.020$. Since both are lower, especially $\sigma_\alpha = 0.004$, it is expected that imposing a common persistence level in the case of DCC may not substantially affect the empirical fit, and this what the LR ratio test result suggests.

The picture from the overall log-likelihood analysis is confirmed by the predictive ability tests for the diagonal models. Compared to the D-OGARCH specifications, D-BEKK produces superior 1-step forecasts with a statistically significant t-statistic equal to 3.49. The D-DCC model outperforms both D-BEKK and D-OGARCH with statistically significant t-statistics equal to 2.66 and 7.09, respectively.

5 Conclusion

This paper advocates a rotation technique for raw returns which leads to easy-to-fit multivariate volatility models via covariance targeting. We discuss the similarities and differences between our approach and the recent orthogonal GARCH models. In particular, while the early contributions to the OGARCH literature assumed, for simplicity, that the estimated orthogonal components are also conditionally uncorrelated, we observe that this is only an approximation since the rotated returns will inherit the conditionally heteroskedastic properties of the unrotated returns. Therefore, we advocate using the popular BEKK and DCC models to study the dynamics of the conditional covariance matrix of the rotated returns. We also discuss a distinct extension of the diagonal BEKK and DCC models, and draw parallels to the OGARCH model of **?**

and the GOGARCH model of **?**.

We show that fitting a diagonal BEKK model to the rotated returns implies a full BEKK specification for the unrotated returns further highlighting the modelling flexibility our approach offers. Estimation and inference is also computationally attractive, thanks to the convenient form of covariance targeting with a long-run identity matrix. Using two-step estimation, we end up estimating only $O(p)$ parameters with numerical optimisation which offers advantages in moderately large dimensions.

Indeed using our approach leads to notable 1-step prediction gains compared to OGARCH and GOGARCH. Capturing the dynamics of the covariances of the rotated returns does improve the prediction of the conditional correlation. Given their flexibility, the DCC suite of models performs best in the 10 dimensional example we study. Interestingly, our newly proposed common persistence model performs quite favourably in comparison to the diagonal model while being more tightly parameterised.

References

	BEKK			OGARCH			GOGARCH			DCC		
	S	D	CP	S	D	CP	S	D	CP	S	D	
Marginal parameters												
Margin (SPY)	–	–	–	–	–	–	–	–	–	(0.077, 0.918)		
Margin (XOM)	–	–	–	–	–	–	–	–	–	(0.084, 0.901)		
Dynamic parameters												
α	0.062	–	–	0.083	–	–	0.072	–	–	0.035	–	
β	0.931	–	–	0.903	–	–	0.921	–	–	0.945	–	
α_{11}	–	0.064	0.077	–	0.088	0.094	–	0.087	0.054	–	0.095	0
α_{22}	–	0.070	0.054	–	0.079	0.066	–	0.074	0.090	–	0.016	0
β_{11}	–	0.932	–	–	0.900	–	–	0.876	–	–	0.904	
β_{22}	–	0.911	–	–	0.899	–	–	0.921	–	–	0.972	
λ	–	–	0.992	–	–	0.986	–	–	0.991	–	–	0
δ	–	–	–	–	–	–	-129.996	-129.999	-129.998	–	–	
LL decomposition												
Margin (SPY)	-3,316	-3,313	-3,313	-3,324	-3,318	-3,321	-3,314	-3,384	-3,330	-3,311	-3,311	-3
Margin (XOM)	-4,030	-4,031	-4,032	-4,026	-4,029	-4,029	-4,029	-4,032	-4,030	-4,026	-4,026	-4
Copula	609	616	611	506	506	509	574	664	597	620	621	
Total LL	-6,737	-6,727	-6,734	-6,844	-6,840	-6,841	-6,769	-6,752	-6,764	-6,717	-6,716	-6

Table 4: Dataset: SPY and XOM daily returns 1/2/2001-31/12/2009. Parameter estimates of the scalar (S), diagonal (D) and common persistence (CP) models. Top panel: Marginal parameter estimates are of the variance targeting GARCH(1,1) models for the DCC margins. Dynamic parameters are estimates of the BEKK, OGARCH and GOGARCH models, and the correlation dynamics for DCC. λ is the common persistence parameter, while δ is the rotation angle in the bivariate GOGARCH model. All parameters are statistically significant at 5 percent. Bottom panel: Log-likelihood decomposition at the estimated parameter values.

	BAC	JPM	IBM	MSFT	XOM	AA	AXP	DD	GE	KO
Eigenvector 1	0.505	0.439	0.182	0.218	0.180	0.360	0.392	0.242	0.292	0.106
Eigenvector 2	-0.584	-0.288	0.177	0.259	0.255	0.582	-0.033	0.223	0.077	0.129

Table 5: Dataset: 10 DJIA stocks daily returns 1/2/2001-31/12/2009. The first two (normalised) eigenvectors correspond to the two largest eigenvalues of the unconditional covariance matrix of the returns.

	BEKK			OGARCH			DCC		
	S	D	CP	S	D	CP	S	D	CP
Dynamic parameters									
α	0.020	–	–	0.045	–	–	0.007	–	–
β	0.978	–	–	0.952	–	–	0.980	–	–
min α_{ii}	–	0.009	0.010	–	0.027	0.025	–	0.004	0.002
max α_{ii}	–	0.054	0.054	–	0.097	0.095	–	0.017	0.015
min β_{ii}	–	0.905	–	–	0.869	–	–	0.932	–
max β_{ii}	–	0.989	–	–	0.967	–	–	0.991	–
λ	–	–	0.998	–	–	0.996	–	–	0.987
LL decomposition									
Margin (BAC)	-4,496	-4,355	-4,373	-4,416	-4,351	-4,361	-4,350	-4,350	-4,350
Margin (JPM)	-4,769	-4,719	-4,734	-4,706	-4,695	-4,700	-4,671	-4,671	-4,671
Margin (IBM)	-4,058	-4,085	-4,092	-4,025	-4,025	-4,023	-4,011	-4,011	-4,011
Margin (MSFT)	-4,449	-4,488	-4,482	-4,438	-4,431	-4,433	-4,424	-4,424	-4,424
Margin (XOM)	-4,090	-4,067	-4,084	-4,040	-4,032	-4,035	-4,026	-4,026	-4,026
Margin (AA)	-5,115	-5,130	-5,132	-5,097	-5,097	-5,096	-5,096	-5,096	-5,096
Margin (AXP)	-4,665	-4,648	-4,652	-4,620	-4,705	-4,654	-4,599	-4,599	-4,599
Margin (DD)	-4,249	-4,310	-4,299	-4,231	-4,247	-4,232	-4,228	-4,228	-4,228
Margin (GE)	-4,291	-4,299	-4,300	-4,263	-4,327	-4,314	-4,257	-4,257	-4,257
Margin (KO)	-3,556	-3,558	-3,562	-3,528	-3,542	-3,542	-3,520	-3,520	-3,520
Copula	4,640	4,860	4,807	3,888	4,040	3,963	4,919	4,946	4,939
Total LL	-39,098	-38,798	-38,904	-39,475	-39,413	-39,426	-38,263	-38,236	-38,244

Table 6: Dataset: 10 DJIA stocks daily returns 1/2/2001-31/12/2009. Parameter estimates of the scalar (S), diagonal (D), and common persistence (CP) models. Top panel: estimates of the dynamic parameters. α and β are the parameters of the scalar models, while $(\alpha_{ii}, \beta_{ii})$, $i = 1, 2$, are those of the diagonal models. For CP, only λ (the common persistence parameter) and α_{ii} are reported. All parameters are statistically significant at the 5 percent level of significance. Lower panel: Log-likelihood decomposition at the estimated parameter values.